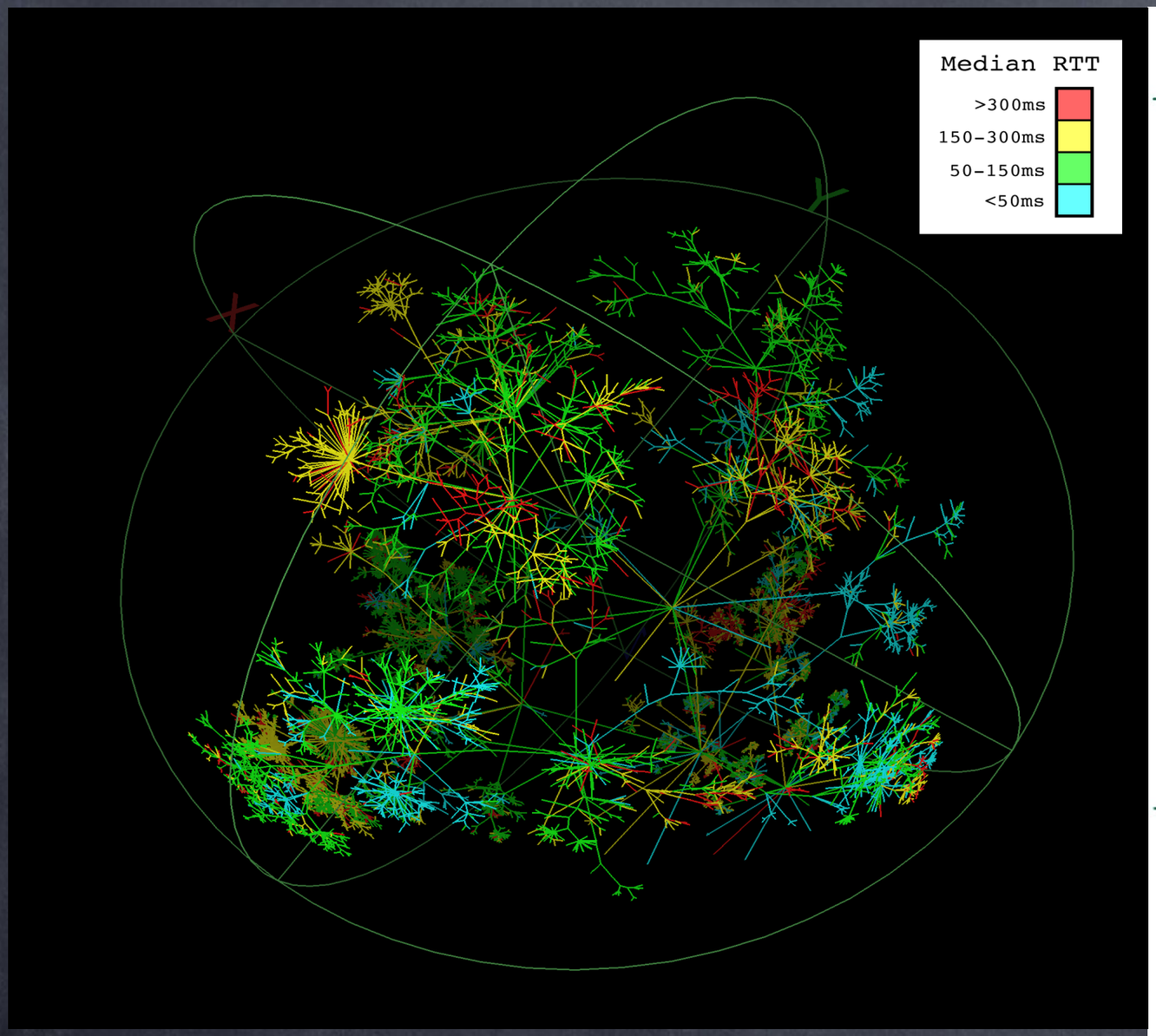
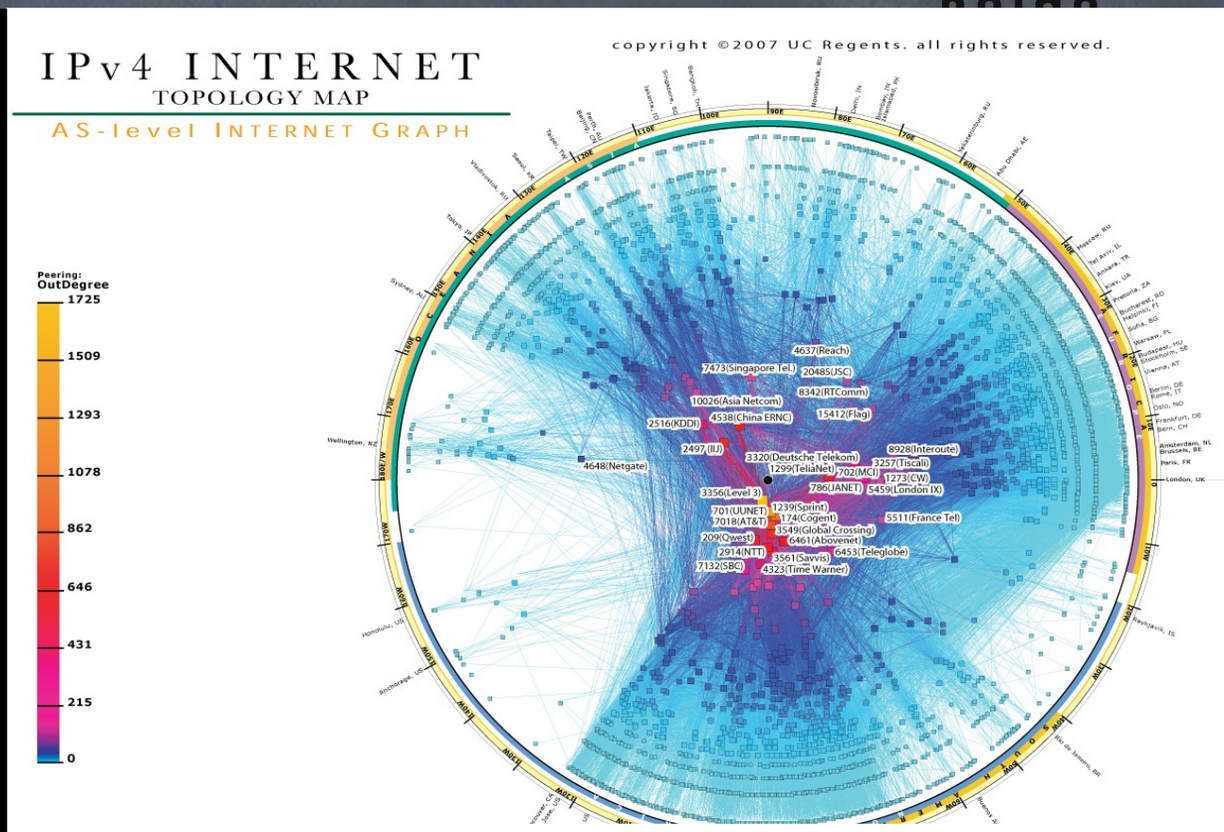
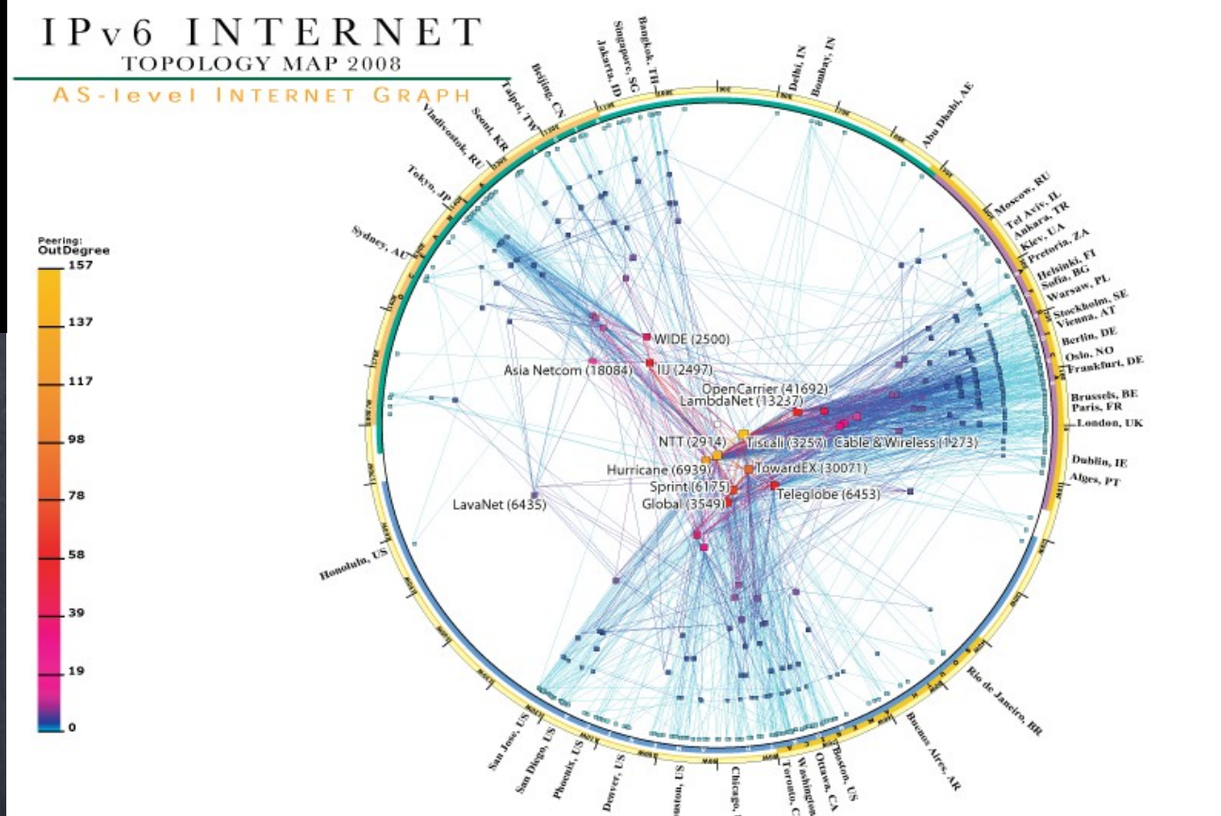
Leveraging the Science and Technology of Internet Mapping for Homeland Security







kc claffy

CAIDA
DHS – PI meeting
SRI Alexandria, VA
10 Sept 2008

Recipe for disaster
(aka “you are here”)

- We now critically depend on the Internet for our professional, personal, and political lives.
- We know little about this information and communication distribution system, e.g, what keeps the system stable or drives it to instability.
- Researchers and policymakers currently analyze an industry in the dark.
- Few data points available suggest a dire picture.
- Agencies charged with infrastructure protection have little situational awareness regarding global dynamics and operational threats

How did we get here?

- Telephone system: 140+ years of history, including regulated data collection requirements (and profits). and a precisely defined system.

- Data networks: 40 years old, ad hoc/hack, tossed to private sector before mature, with no govt support for research or metrics (or profit), ill-defined system.

- Current academic projects either lack sustainability (iplane) or ability to dedicate resources (PlanetLab)

- War: the best motivation so far for investing in situational awareness of critical infrastructure

Approach: a new architecture: ark

- CAIDA's new measurement infrastructure
- Build on decade of achievements, from SIGCOMM to MOMA
- Launch 12 Sept 2007
- 28 active probers
- 5 are IPv6-capable
  - collaborators can run vetted measurements on security-hardened platform
  - general public can perform restricted measurements
  - support for meta-data mgt, analysis, and infoviz

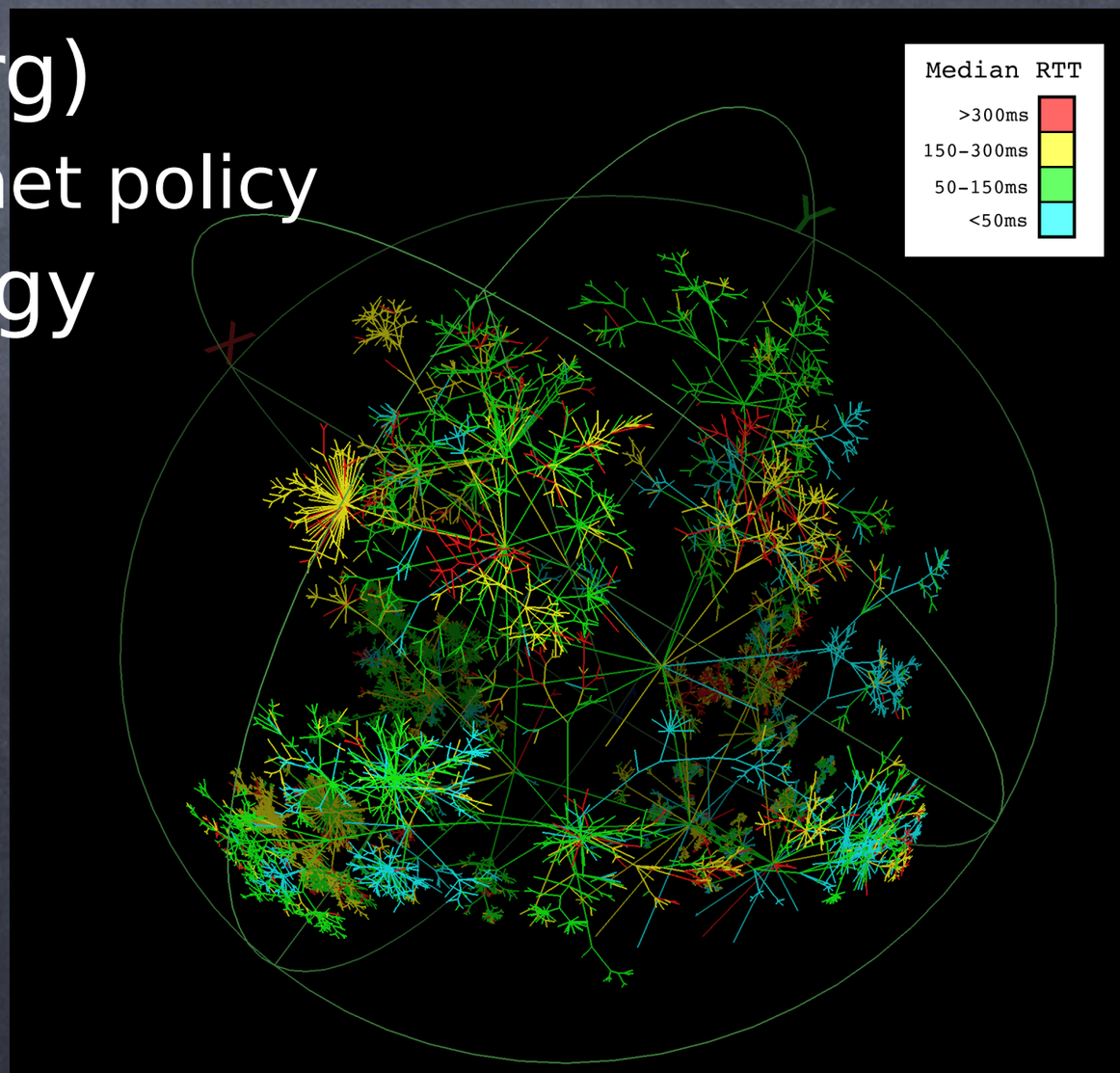
Connect with SA requirements

Benefits

- Improve critical national capabilities:
  - situational awareness for homeland security purposes
  - topology mapping
  - internet measurement, analysis and inference techniques
  - empirical basis for federal communications policy
- Address network science crisis
  - scalability in system management, monitor deployment, measurement efficiency, resource utilization
  - flexibility in measurement method, scheduling, data collection
  - let researchers spend less time on non-research

Profound insights enabled

- Incongruity between topology and routing system
  - topology evolving away from what routing system needs
  - radical implication for future of the Internet (IP)
- Concentration of ISP ownership (as-rank.caida.org)
  - Inform communications, Internet policy
- Incongruity between topology and routing data
  - still no guaranteed way to capture Internet topology
  - but some methods are better than others, e.g., ICMP



Internet Mapping: Simple Example

- Need: What probing method discovers most topology?
  - Do per-flow load balancers implement different forwarding policies for TCP and UDP? Need experiment!
- Approach: Archipelago measurement platform
  - Matthew Luckie, Young Hyun, and Bradley Huffaker, “Traceroute Probe Method and Forward IP Path Inference”, IMC 2008.
    - ICMP-based traceroute methods tend to successfully reach more destinations, as well as collect evidence of a greater number of AS links.
    - UDP-based methods infer the most IP links, despite reaching the fewest destinations.

Internet Mapping: Simple Example (cont)

- Benefits:
  - Ease of experiment design, implementation, and coordination.
  - Dedicated resources (monitors).
  - No restrictive intellectual property.
  - Multiple levels of trust and access
- Competition:
  - Other platforms:
    - do not provide dedicated resources.
    - cannot guarantee the veracity of the collected data.
    - lack fine granularity access control
  - Other data collected on these platforms suffer the constraints of the underlying platform.

Approach

- Integrate 6 strategic measurement and analysis capabilities:
  - new architecture for continuous topology measurements,
  - IP alias resolution techniques,
  - dual router- and AS-level graphs,
  - AS taxonomy and relationships,
  - geolocation of IP resources, and
  - graph visualization.

Competition

- PlanetLab (http://www.planet-lab.org/)
  - research and resource constraints (non-dedicated)
- iPlane (http://iplane.cs.washington.edu/)
  - runs on PlanetLab
- DIMES (http://www.netdimes.org/)
  - no control over monitors (run on end-user h/w)
  - cannot trust data

- Maybe more importantly, barriers to success (of measurement)
  - Economics
    - cost of keeping pace with backbone link technology
  - Ownership
    - proprietary networks with disincentive to share data
  - Trust
    - privacy issues and methods of protecting personal information

Why is Internet mapping worthwhile?

- Need: situational awareness: to provide richly annotated maps of the Internet to support better understanding of this critical infrastructure for national security and communications policy needs.
- **A**pproach: integrate 6 strategic measurement and analysis capabilities.
- **B**enefits:
  - improved situational awareness for homeland security purposes
  - improved topology mapping
  - improved internet measurement and property inference techniques
  - improved network analysis techniques
  - improved empirical basis for federal communications policy
  - improved science of the Internet
- **C**ompetition (but not really):
  - PlanetLab (http://www.planet-lab.org/)
  - iPlane (http://iplane.cs.washington.edu/)
  - DIMES (http://www.netdimes.org/)

Nugget of CAIDA's Internet Mapping

- **Archipelago** provides a unique enabling infrastructure, featuring the Miranda tuple space, that allows researchers to quickly design, implement, and easily coordinate the execution of experiments across a widely distributed set of dedicated resources (monitors). Ark coordination facilities also enable ease of data transfer, indexing, and archival.

2008 Technical accomplishments

- 28 monitors now active
- raw IPv4 topology data (July deliverable)
  - 200M paths, served thru PREDICT and data.caida.org
- Converted as-rank.caida.org to use Ark data
- Probing method comparison: IMC2008 paper
- Incorporated more sources of BGP data (RV and 17 RIPE srcs) into IP-->AS mapping
- Written summary of using annotations in dual-level graph (available upon request)
- ***iffinder*** experiment with 24 cycles of Ark data
  - Analyzed results: 3% reduction, will feed into APAR
- Modified APAR code for scalability

Approach: IP Alias Resolution

- collapse IPs into the same router
- all techniques have strengths and weaknesses, so we combine them to get the best results
- our plan:
  - run **iffinder** on Routed /24 data
  - run **APAR** using iffinder results as seed
  - run **Ally** on final set of aliases, as validation

Approach: IP Alias Resolution (cont.)

- how much topology data should we examine?
  - about time period (window), not quantity
    - last month, 3 months, or year of traces?
  - window must be large enough
    - include topology traversed infrequently or irregularly
    - in Routed /24 Topology dataset, only **one** monitor probes each /24 per cycle
  - window should not be too large
    - may include topology that no longer exists
    - will increase amount and difficulty of processing

Approach: Dual Router- and AS-level Graphs

- Map traceroute data to AS-level
  - conceptually simple, well known
  - use Route Views BGP tables
  - discard and filter ~5% of links in the process
    - AS sets, multi-origin & private ASes, indirect links

- Two distinct topologies: AS and router- level
- Need to merge into a dual graph
  - assign routers to ASes

- Will evaluate multiple techniques
  - *dK-series*, CAIDA powerful methodology for topo analysis

Approach: AS Taxonomy and Relationships

- CAIDA has developed an AS classification scheme resulting in the most veracious Internet AS taxonomy to date.
- We classify 95.3% of ASes with an expected accuracy of 78.1%. We annotate each AS with:

١) the organization description record,
٢) the number of inferred customers,
٣) the number of inferred providers,
٤) the number of inferred peers,
٥) the number of advertised IP prefixes, and
٦) the equivalent number of /24 prefixes covering all the advertised IP space.

Approach: AS Taxonomy and Relationships (cont)

- We release to the community the Autonomous System Taxonomy Repository as well as:
  ١) the AS taxonomy information and
  ٢) the set of AS attributes we used to classify ASes.
- Improve and enrich AS-ranking suite
  - based on AS relationship heuristics
  - will benefit from better measurement data
- Telco hotel data integration (if available)

Approach: Geolocation of IP Resources

- CAIDA currently makes use of Digital Envoy's NetAcuity IP address geolocation services.
- We would like to conduct geolocation “cookoff” to find best of breed tools for geolocation.
- CAIDA has domain experience gained through development, maintenance, and support of open source tool, NetGeo. Still used by many but no longer supported.

Schedule, Planned activities

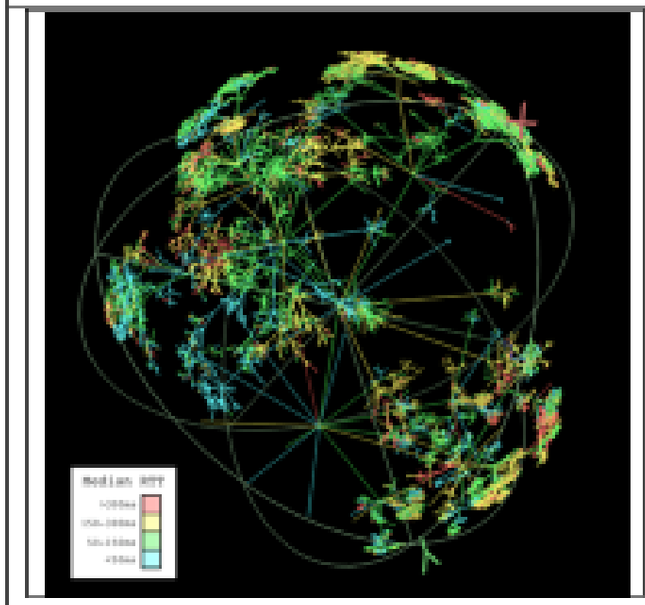
- 1-2 monitors/month
- IPv4, IPv6 topology data
- Characterize load-balancing behavior
- Try other approaches to dual-graph construction
- Continue alias resolution study, derive recommendations (inc. another iffinder run)
- Ask friendly providers (e.g., I2) for validation of topology inferences (ground truth)
- Better viz with walrus
- Early 2009: workshop on utility of infrastructure

Tech transition plan

- Software tools publicly available (UCSD or GPL license)
- Early 2009: workshop on utility of infrastructure tied to PREDICT workshop on utility of data available from other operational infrastructure.

BAA Number: Cyber Security BAA 07-09
Title: Science and Technology of Internet Topology Mapping

Offeror Name: Kimberly Claffy
Date: 06/26/07



Walrus visualizations of round-trip time measurements made by CAIDA's macroscopic Internet topology monitor located in Herndon, VA, USA.

Internet Topology Mapping:

1. Operational infrastructure to support continuous Internet topology mapping.
2. Periodic active probing of 100% of BGP prefixes announced in publicly available routing tables.
3. ISP relationship inference with accuracy up to 98%.
4. Topologies at the router and AS granularity annotated with AS relationships, AS types, geolocations, latencies, etc.
5. Empirically grounded quantified understanding of robustness, reliability, scalability and other characteristics of the Internet topology as critical infrastructure.
6. Improved annotated topology maps will enhance modeling and monitoring capabilities to help identify threats and predict cascading impacts of damage scenarios.
7. Visualization capabilities will provide powerful interface for use by DHS and other national security personnel.

Technical Approach:

1. Expand current deployment of new distributed platform for continuous measurement of Internet topology, performance, state, and other characteristics.
2. Use and improve IP alias resolution techniques to identify common routers to which IP interfaces belong.
3. Further test and improve performance of software to convert IP technology data into router-level and AS-level graphs.
4. Utilize CAIDA's AS relationship and AS taxonomy inference techniques and data infrastructure to annotate AS graphs with AS types and relationships.
5. Apply and evaluate publicly available geolocation tools for use in annotating topologies with geographic data.
6. Use CAIDA's or other visualization capabilities to depict structure and vulnerability-related characteristics of observed annotated Internet topologies.

Schedule, Deliverables, Contact Info:

1. Current: new active measurement architecture: design complete; prototype implementation being tested.
2. Year 1:
   a. establish on-going IPv4 topology measurements using the new infrastructure;
   b. release software for calculation and exhaustive analysis of topology characteristics.
3. Year 2:
   a. weekly updates of router topology with IP aliases resolved using best available techniques;
   b. weekly updates of AS/router graphs annotated with inferred AS relationships and types.
4. Year 3:
   a. topology annotated with latencies and geolocations;
   b. annotated AS/router topology visualizations.
5. POC: Jennifer Ford, UCSD Contracts&Grants, 9500 Gilman Dr. MC 0934, La Jolla, CA 92093-0934 Fax : (858) 534-0280

Links

- Archipelago (Ark) network measurement platform http://www.caida.org/projects/ark/
- Autonomous System Taxonomy Repository http://www.caida.org/data/active/as_taxonomy/
- Internet Measurement Conference http://www.imconf.net/imc-2008/